Fisheries and Oceans Canada | Pêches et Océans Canada
Science | Sciences

Canadian Science Advisory Secretariat (CSAS)

Research Document 2013/076

Maritimes Region

Within- and among-population genetic variation in Eastern Cape Breton Atlantic Salmon and the prioritization of populations for conservation (*Salmo salar* L.)

P. O'Reilly, S. Rafferty, and A.J.F. Gibson

Fisheries and Oceans Canada
Science Branch, Maritimes Region
P.O. Box 1006, Dartmouth, Nova Scotia
Canada B2Y 4A2

December 2013

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research documents are produced in the official language in which they are provided to the Secretariat.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

© Her Majesty the Queen in Right of Canada, 2013
ISSN 1919-5044

Correct citation for this publication:

O'Reilly, P., Rafferty, S., and Gibson, A.J.F. 2013. Within- and among-population genetic variation in Eastern Cape Breton Atlantic salmon and the prioritization of populations for conservation *(Salmo salar* L.). DFO Can. Sci. Advis. Sec. Res. Doc. 2013/076. vi + 25 p.

TABLE OF CONTENTS

ABSTRACT iv
RÉSUMÉ v
INTRODUCTION 1
METHODS 2
RESULTS AND DISCUSSION 5
Within-population genetic variation 5
Among-population genetic structuring 8
Implications of observed within- and among-population genetic variation for the prioritization of conservation efforts for East Cape Breton Atlantic salmon 10
Caveats and limitations 11
ACKNOWLEDGEMENTS 12
REFERENCES 12
TABLES 16
FIGURES 19

ABSTRACT

Genetic variation was surveyed at 15 microsatellite loci from 798 Atlantic salmon (*Salmo salar* L.) juveniles collected from seven rivers of the Eastern Cape Breton Designatable Unit, and several reference rivers from elsewhere in Atlantic Canada, including two from the west coast of Cape Breton. In general, levels of within-population genetic variation (allele richness and expected heterozygosity) were slightly lower in East Cape Breton salmon relative to salmon from the large reference populations analyzed, but differences were not significant. Within East Cape Breton, population genetic variation was slightly reduced in salmon from the far east coast of Cape Breton Island (Grand and Inhabitants rivers) compared to salmon from North Central Cape Breton (North Aspy, North (Victoria), Baddeck and Middle rivers), but again these differences were modest and not statistically significant. The Indian Brook sample collection (the only one obtained from the Bras d'Or Lake) exhibited the lowest level of allele richness of any originating from East Cape Breton, approaching levels observed in the sample collection obtained from the small (Stewiacke River) reference population included here. Although overall, reductions in levels of genetic variation were modest in East Cape Breton sample collections compared to reference collections, many also exhibited multiple pairs of loci that appear to be in linkage disequilibrium; a subset of these same sample collections also exhibited multiple loci that deviate from Hardy-Weinberg expectations. Taken together, these data indicate that several East Cape Breton river populations (North Aspy, Indian Brook, Grand and Inhabitants) may have experienced recent population bottlenecks.

Levels of genetic structuring (F_{ST}) between sample collections obtained from different East Cape Breton rivers varied from 0.0035-0.0363 ($\overline{X}$ =0.0193), and all pairwise comparisons were significant following adjustments for multiple tests. Sample collections from North Central Cape Breton, with the exception of that obtained from the North (Victoria) River, exhibited very little genetic differentiation (F_{ST}=0.0035-0.0.0083, $\overline{X}$ F_{ST} =0.0.00653), and cluster closely together in phylogenetic and factorial correspondence analyses. These same three sample collections also exhibited very little differentiation from that obtained from the nearby Margaree River of the Gaspé-Southern Gulf of St. Lawrence Designatable Unit ($\overline{X}$ F_{ST} =0.0086), and this similarity was again reflected in both phylogenetic and factorial correspondence analyses. The Indian Brook sample collection was well differentiated from all others analyzed, and represents a second major grouping of East Cape Breton salmon surveyed here. Inhabitants and Grand sample collections were relatively differentiated from each other, and from all others from East Cape Breton in multiple analyses, though the former showed some similarity to the sample collection obtained from the Mabou River of the Gaspé-Southern Gulf of St. Lawrence Designatable Unit, and the latter similarity to the sample collection obtained from the St. Mary's River of the Southern Upland Designatable Unit. These two sample collections (Grand and Inhabitants) may represent two additional groupings of East Cape Breton Atlantic salmon. Possible indications of within-population structuring was observed in North (Victoria) sample collection, suggesting the presence of a fifth group of East Cape Breton Atlantic salmon, though additional analyses of further samples from this location are required to substantiate these latter findings. Patterns of within- and among-population genetic variation observed here may help inform efforts to prioritize Atlantic salmon populations from the East Cape Breton Designatable Unit for conservation.

Variation génétique au sein des populations et entre les populations du saumon de l'Atlantique (*Salmo salar* L.) de l'est du Cap-Breton et établissement des populations prioritaires aux fins de conservation

RÉSUMÉ

La variation génétique a été examinée sur 15 loci microsatellitaires de 798 saumons de l'Atlantique (*Salmo salar* L.) juvéniles de sept rivières de l'unité désignable de l'est du Cap-Breton et plusieurs rivières de référence ailleurs au Canada atlantique, y compris deux rivières de la côte ouest du Cap-Breton. En général, les niveaux de variation génétique au sein des populations (nombre d'allèles et hétérozygotie prévue) étaient légèrement plus bas chez les saumons de l'est du Cap-Breton que chez les saumons des grandes populations de référence analysées, mais les différences n'étaient pas importantes. Dans l'est du Cap-Breton, la variation génétique au sein des populations était légèrement plus basse chez les saumons de l'extrémité est du Cap-Breton (rivières Grand et Inhabitants) que chez les saumons du centre nord du Cap-Breton (rivières North Aspy, North [Victoria], Baddeck et Middle), mais, encore une fois, ces différences étaient légères et n'étaient pas statistiquement significatives. Les échantillons prélevés dans le ruisseau Indian (la seule collecte d'échantillons du lac Bras d'Or) présentaient le plus faible nombre d'allèles parmi toutes les populations provenant de l'est du Cap-Breton; les niveaux étaient près de ceux observés chez les échantillons obtenus de la petite population de référence (rivière Stewiacke) dans le cadre de la présente étude. Dans l'ensemble, les réductions des niveaux de variation génétique étaient légères chez les échantillons prélevés dans l'est du Cap-Breton par rapport aux populations de référence. Toutefois, de nombreux individus avaient plusieurs paires de loci dont les liens semblaient être en déséquilibre, et un sous-ensemble de ces mêmes échantillons avait aussi plusieurs loci montrant des écarts par rapport aux prédictions de Hardy-Weinberg. Ensemble, ces données indiquent que plusieurs populations des rivières de l'est du Cap-Breton (North Aspy, Indian Brook, Grand et Inhabitants) pourraient avoir connu une récente pénurie de population.

Les niveaux de structuration génétique (F_{ST}) entre les échantillons de différentes rivières de l'est du Cap-Breton variaient de 0,0035 à 0,0363 ($\overline{X}$ =0,0193), et toutes les comparaisons par paire étaient importantes à la suite de modifications pour plusieurs tests. Les échantillons du centre nord du Cap-Breton, à l'exception de ceux de la rivière North (Victoria), présentaient une très faible différenciation génétique (F_{ST}=0.0035-0.0.0083, $\overline{X}$ F_{ST} =0.0.00653), et étaient regroupés dans les analyses phylogénétiques et factorielles de correspondance. Ces trois mêmes échantillons présentaient aussi une très faible différenciation par rapport aux échantillons d'une rivière située à proximité, à savoir la rivière Margaree, de l'unité désignable Gaspésie–sud du golfe du Saint-Laurent ($\overline{X}$ F_{ST} =0,0086), et cette ressemblance était encore une fois reflétée dans les analyses phylogénétiques et factorielles de correspondance. Les échantillons du ruisseau Indian présentaient d'importantes différences par rapport aux autres échantillons analysés; ils représentent un deuxième important groupe de l'est du Cap-Breton examiné au cours de la présente étude. Les échantillons des rivières Inhabitants et Grand présentaient des différences entre eux ainsi que par rapport aux autres analyses de l'est du Cap-Breton. Cependant, le premier affichait plus de ressemblances avec les échantillons de la rivière Mabou, de l'unité désignable Gaspésie–sud du golfe du Saint-Laurent, tandis que le deuxième affichait des ressemblances avec les échantillons de la rivière St. Mary's, de l'unité désignable des hautes terres du Sud. Ces deux échantillons (Grand et Inhabitants) pourraient représenter deux autres groupes du saumon de l'Atlantique de l'est du Cap-Breton. Des indicateurs possibles de la structuration au sein des populations ont été observés dans les échantillons de la rivière North (Victoria), ce qui laisse entendre la présence d'un cinquième groupe du saumon

de l'Atlantique de l'est du Cap-Breton. Toutefois, des analyses additionnelles d'autres échantillons de cet emplacement sont nécessaires afin de corroborer cette conclusion. Les tendances en matière de variation génétique au sein des populations et entre les populations observées au cours de la présente étude pourraient aider à orienter les efforts visant à établir les populations de saumon de l'Atlantique de l'unité désignable de l'est du Cap-Breton prioritaires aux fins de conservation.

INTRODUCTION

East Cape Breton (ECB) Atlantic salmon include those that, as juveniles, inhabit streams and rivers that drain into Atlantic waters along the northeast, southeast, and south coasts of Cape Breton Island, as well as those inhabiting streams and rivers that empty into the Bras d'Or Lakes. Salmon of ECB have previously been identified as distinct from Atlantic salmon elsewhere in Eastern Canada (O'Reilly 2006; DFO and MNRF 2008) and because of past and recent declines, were designated for listing as endangered in 2010 (COSEWIC 2010). Just as the maintenance of major groupings (Designatable Units, DUs) of Atlantic salmon are necessary for the conservation of the species as a whole (DFO and MNRF 2008), identification and protection of biodiversity within DUs may be important for the conservation of the DU itself.

Cape Breton Island includes several geographic characteristics that may be expected to contribute to restricted gene flow, local adaptation, and the development of important biodiversity within this group of Atlantic salmon. While many rivers along the east coast of the island drain almost directly into the Atlantic Ocean, rivers in the central part of the island empty into the Bras d'Or Lakes, an inland sea of brackish water connected to the Atlantic by several kilometres of narrow channels (discussed further below) (Figure 1). Up until the construction of a canal in 1869, water in the most internal of the Bras d'Or Lakes, Bras d'Or Lake itself, had first to pass through Barra Strait, then Great Bras d'Or Lake, then one of several narrow channels before meeting the Atlantic Ocean (Figure 1). This complex coastline, narrow channels and brackish water may have influenced the likelihood of non-local strays enccuntering and ascending rivers of the Bras d'Or lakes, particularly those of the internal Bras d'Or Lake.

The main body of Cape Breton Island, that which lies north of the Bras d'Or Lake, is dominated by the Cape Breton Highlands plateau, an extension of the Appalachian Mountains. Much of plateau lies between 300 and 500 metres above the sea level. Rivers whose watersheds are contained within these Highlands might be expected to exhibit physical and biological characteristics quite different from those in surrounding Lowlands to the west. For example, Highland rivers are often characterized by steep gradients and swift currents. Water velocity of lower order streams has been shown to be associated with variation in body morphology, migration timing and fin size in juvenile North American Atlantic salmon (Riddell and Leggett 1981). High and lowland habitats may also be expected to exhibit different thermal regimes. This may be particularly important, as recent research (discussed below) is indicating that spring and summer stream temperature experienced by juveniles may be important in influencing gene flow and the development of local adaptation in Atlantic salmon. Although many factors are associated with stream temperature, including extent of overhanging cover and stream source, water temperatures are closely associated with air temperature (Mohseni and Stefan 1999), which in turn is strongly influenced by elevation. Given the increased elevation of the headwaters in the Highlands (see above) compared to some lowland rivers (50 to 150 metres) summer water temperatures in headwater streams can be expected to vary moderately across Cape Breton, despite the limited latitudinal range (45º28" to 47º2") encompassed by the island. Indeed, rivers of ECB, including several analyzed here, do appear to vary markedly in terms of three thermal indices thought to be related to general salmonid health, including mean summer (June 15 to Sept 5) temperature, warmest daily averages, and number of days above 20 degrees Celsius (MacMillan *et al.* 2005).

There is now a considerable body of evidence for the existence of local adaptation in Atlantic salmon associated directly or indirectly with temperature differences among rivers. Verspoor and Jordan (1989) observed a south to north cline in the frequency of a particular malic acid allozyme variant that was paralleled by a low to high elevation cline in the frequency of this same allele; this has been cited as some of the strongest evidence for local adaptation in

salmon (Taylor 1991). Nicieza *et al.* (1994) reported countergradient variation in passage time and digestive rates of Atlantic salmon inhabiting cool high altitude and warmer low altitude environments. Common garden experiments involving salmon from these same disjunct populations demonstrate a large genetic component to the observed differences; together, these findings suggest that this variation may be adaptive, allowing for more rapid growth in cooler high altitude environments (Nicieza *et al.* 1994). More recently, Dionne *et al.* (2007) reported a south to north cline in levels of diversity in the peptide binding region of the Major Histocompatabilty Class II locus in Atlantic salmon from Eastern Canada. The observation that pathogen diversity also increased with latitude in these same rivers led the authors to suggest the molecular genetic variation observed may be an adaptation to bacterial diversity in the populations surveyed. Finally, landscape analyses of neutral microsatellite variation in Atlantic salmon from 51 rivers along the Atlantic coast of Quebec, Labrador, and Northern New Brunswick indicated that gene flow and adaptation to thermal regimes (as opposed to the multiple other possible factors examined in the study) explained most of the genetic structure observed across the populations surveyed (Dionne *et al.* 2008).

Examination of patterns of molecular genetic variation can be useful in identifying and prioritizing remaining within- species biodiversity for conservation actions. For example, analysis of DNA variation has helped identify major ancestral lineages in salmonids that were not otherwise apparent (Utter *et al.* 1993; Verspoor *et al.* 2002). Additionally, analyses of patterns and extent of genetic structuring among samples from different locations can provide insight into the extent of recent and ongoing gene flow. This information is important in inferring the potential for adaptive differences to have developed between salmon from different rivers or regions, because genetically based adaptive differentiation can only accrue in the absence of large amounts of gene flow (Waples 1991). Assessments of levels of within- and among-population genetic variation have also been used directly to prioritize populations for conservation efforts (Petit *et al.* 1998) with, all else being equal, more weight given to those exhibiting higher levels of within-population variation, and to those that are more genetically divergent from others (discussed further below). This increased importance of more genetically diverse populations may reflect both a) potential increased likelihood of persistence of a more genetically variable population over less variable population (Saccheri *et al.* 1998) and hence the ability of a population to contribute demographically to the species through time, and b) the potential increased adaptability of a population exhibiting higher levels of within-population genetic variation in the face of future environmental change (Nevo 1978).

Here, genetic variation is surveyed at a large number of nuclear microsatellite loci (15) across 798 individuals obtained from 13 rivers, including seven from ECB, two from nearby rivers on the West Coast of Cape Breton (WCCB) in the neighbouring Gaspé-Southern Gulf of St. Lawrence (GSG) DU, one from a nearby river in the neighboring Southern Upland (SU) DU, and three more distant locations that vary in size and levels of within-population genetic variation (Figure 1; Table 1). The objectives of these analyses are (1) to report and interpret levels of within-population genetic variation in ECB populations surveyed, (2) to quantify the extent and pattern of present-day among-population genetic structuring within the ECB DU, and (3) to prioritize populations for conservation measures based on 1 and 2 above.

METHODS

River and sample collection information: Sample collections were obtained from seven ECB rivers covering much of the DU, beginning with North River Aspy, which drains almost directly into Atlantic waters at the northern tip of Cape Breton Island (Figure 1). Collections obtained from North River Aspy in 2006 and 2007 were combined in order to increase the overall sample size to 44; the resulting sample collection was designated NRA0607 (Table 1). Continuing

south along the central axis of the island is North River of Victoria County, which empties into the long and narrow St. Ann's Bay. This collection was obtained in 2006, and is designated NRV06. Further south along the central axis of the Island are Baddeck and Middle rivers, which empty into Nyanza Bay of Great Bras d'Or Lake. Sample collections from these two rivers were obtained in 2010 and 2006, and are designated BAD10 and MDV06, respectively. Great Bras d'Or Lake is connected to the Atlantic waters north of Cape Breton Island via the Great Bras d'Or Channel running along the north shore of Boularerie Island, and St. Andrews and Little Bras d'Or channels running along the south shore of Boularerie Island. A single river emptying into Bras d'Or Lake, Indian Brook, was also sampled. Collections obtained from Indian Brook in 2006 and 2007 were also combined in order to increase the overall sample size to 52, and the resulting sample collection designated IND0607. Historically, salmon travelling to and from Indian Brook had to pass through Barra Strait (approximately half a kilometre wide at its narrowest point), then Great Bras d'Or Lake and the associated channels, to move between the Atlantic Ocean and their natal river. In 1869, however, St. Peter's Canal was constructed at the extreme south end of the Bras d'Or Lake, connecting the water body directly to Atlantic waters off the South Coast of Cape Breton Island. Although the extent to which Atlantic salmon of the Bras d'Or Lake system currently use St. Peter's Canal on route to and from the Atlantic Ocean is unknown, surface and sub-surface flows between Bras d'Or Lake and the Atlantic southeast of Cape Breton Island (<100 m^3/s) are a small fraction of that moving between Great Bras d'Or and the Atlantic northeast of the Island (750-1500 m^3/s), and between Great Bras d'Or and Bras d'Or lakes (750-1500 m^3/s) (Parker *et al.* 2007). Two rivers were also sampled along the Southeast Coast of Cape Breton (SECB), Grand which empties almost directly into the Atlantic Ocean, and further south, Inhabitants, which empties into Inhabitants Bay, then the Strait of Canso, and finally the Atlantic Ocean south of Cape Breton Island. Sample collections from these two rivers were both obtained in 2010, and are designated GRA10 and INH10, respectively.

The watersheds occupied by these seven rivers vary ecologically. Greater than 50 percent of the watersheds of North River Aspy, North River (Victoria), Baddeck and Middle rivers are at higher elevations (greater than 250 metres), in the Ecodistrict defined as Cape Breton Highlands; see Neily *et al.* (2003) for descriptions of this and other Ecodistricts listed below. A smaller portion of the North River Aspy watershed falls within the Victoria Lowlands and Cape Breton taiga Ecodistricts, and very small portions of the North River (Victoria) watershed is comprised of Bras d'Or Lowlands and Cape Breton Hills Ecodistricts. A moderate percentage of the Baddeck watershed is contained within Bras d'Or Lowlands and Cape Breton Hills Ecodistricts, and a small portion of the Middle watershed is contained within the Inverness Lowlands Ecodistrict. The Indian Brook watershed is almost entirely contained within the Cape Breton Hills Ecodistrict, much of which is at an elevation of 150 and 200 metres above sea level. Large portions of the Inhabitants watershed are contained within the Cape Breton Hill Ecodistrict (150-200 metres in elevation) and the Bras d'Or Lowlands Ecodistrict (below 100 metres). Most of the Grand River watershed is contained within the Bras d'Or Lowlands and Cape Breton Coastal Ecodistricts, and is below 100 metres in elevation.

Also included in this study are two reference sample collections obtained from rivers from the WCCB, Margaree and Mabou, located in the neighboring GSG DU, and one reference population from the St. Mary's River within the neighboring SU DU. The Margaree collection was obtained in 2001, and is designated MRG01, and Mabou collection in 2006, and is designated MAB06. The St. Mary's River collection was obtained in 2007, and is designated SME07. The primary purpose of these reference sample collections is to inform the interpretation of patterns of among-population variation observed within ECB, although they are also used in comparisons of levels of within-population genetic variation. Two large reference populations from New Brunswick (NB), one from the Nashwaak River of the outer Bay of Fundy

DU, and the other from the Kedgwick River (a tributary of the Restigouche River) located within the GSG DU, were sampled to analyze and compare levels of within-population genetic variation observed in ECB sample collections. The Nashwaak River collection was obtained in the year 2000, and was designated NSH00, while the Kedgwick River collection was obtained in 2003, and is designated RKR03. Samples from the small Stewiacke River of the inner Bay of Fundy (iBoF) DU were included, also for the purpose of comparing levels of within-population genetic variation. The Stewiacke River collection was obtained in 2001, and was designated STW01. With the exception of those obtained from North River Aspy and Indian Brook (discussed above), sample collections were obtained in a single year. All sample collections were comprised of parr obtained by electrofishing multiple dispersed sites within a river. River names, sample sizes, sample collection codes, geographic coordinates, and associated DUs and Salmon Fishing Areas (SFAs), are given in Table 1.

Laboratory analyses: Fin clips were collected and stored in 1.5 ml microcentrifuge tubes containing 1 ml of ethanol. DNA was extracted and purified using Qiagen's 96-well DNeasy Blood and Tissue kits following the manufacturer's specifications. Polymerase Chain Reaction (PCR) amplifications were carried out for each locus separately in 10 µl volumes containing between 1-100 nanograms of template DNA, 0.2 mM each dNTP, 0.1 µM labelled and unlabelled primers, 1X KCl buffer (10 mM Tris HCl, 50 mM KCl, 0.08% Nonidet P40), 0.5 units of Taq DNA polymerase supplied by MBI Fermentas and 2.5 mM $MgCl_2$. Primer sequences for loci *Ssa*85, *Ssa*197 and *Ssa*202 are given in O'Reilly *et al.* (1996); SS*sp*2201, SS*sp*2210, SS*sp*2215, SS*sp*2216, SS*sp*1G7 and SS*sp*1605 are given in Paterson *et al.* (2004); *Ssa*D58, *Ssa*D144, *Ssa*D71, and *Ssa*D486 in King *et al.* (2005); and *Ssos*L417 in Slettan *et al.* (1995). Primers for the locus *Ssa*D85 are unpublished, but are CTTTGGCTGTTTCAGGTATGAC and CACTGCTCTACAACAGAAGTCTC (T. King, Genbank Accession AF525213). Thermal cycling conditions were as follows: (94°C for 3 min)X1, (94°C for 45 sec, 58°C for 30 sec, 74°C for 1 min)X9, and (94°C for 30 sec, 58°C for 30 sec, and 72°C for 1 min)X 27, followed by a 30-min extension step at 72°C. PCR amplification products from individual loci were size-fractionated without purification or products from multiple loci were combined in various groups, and salt, unincorporated dNTP's and non-labeled primers removed using either Qiagen's 96-well MinElute plates (as specified by the manufacturer) or using ethanol precipitation methods prior to size fractionation. Alleles were size-fractionated using denaturing polyacrylamide electrophoresis, and detected using either an MJ Research Base station or an Applied Biosystems 3130XL. Size determinations of fragments analyzed in different batches, on different days and on different platforms, were cross-standardized by including 2 of 10 individuals with known genotypes in each batch of samples analyzed; these two individuals varied across batches, providing an "internal label" in addition to an external written label to identify batches of sampled analyzed. Ten samples from each batch of 84 were analyzed twice to identify potential sample placement errors, strip inversions, plate inversions and other laboratory errors.

Estimation of within-population genetic variation: Several different measures of within-population genetic variation were estimated for each sample collection and each locus, and values averaged across all 15 loci. The *observed number of alleles (NaO)*, was simply the number of different alleles observed in a given sample collection, with no attempts to control for sample-size effects. To permit comparisons of numbers of alleles observed across sample collections of varying size, the standardized number of alleles or *Allele richness (AR)* was estimated using *FSTAT* version 2.9.3.2 (Goudet 1995, 2001), which is based on the rarefaction procedure of Hurlbert (1971). In this approach, estimates of the expected number of different alleles for each population are made by repeated sampling of 2N genes, where N is the smallest sample size of diploid genotypes present among the sample collections under study (39 in this study). The *observed heterozygosity (ObH)*, was simply the proportion of genotypes exhibiting

two different alleles. *Gene diversity (GD)*, the likelihood that two alleles randomly drawn from a sample are different, was also estimated using *FSTAT*. The extent of non-random mating, F_{IS}, (*f* from Weir and Cockerham 1984), approximately equal to (*Hs-Ho/Hs*), where *Hs=GD* here, and significance of departures from zero were estimated using *Genetix* version 4.02 (Belkhir *et al.* 2000). Tests for linkage disequilibrium, or whether genotypes at one locus were independent of genotypes at another, were carried for all pairs of loci and for all sample collections, using option 4.2.1 of *Genepop* version 4.0 (Rousset 2008).

Analysis of levels and patterns of among-population genetic variation: An analog of $F_{ST}(\theta)$ (Weir and Cockerham 1984), and significance of deviations from zero (homogeneity), were estimated for all pairs of sample collections using permutation methods (1000 samples) implemented in the program *Genetix* version 4.02 (Belkhir *et al.* 2000). The program *Populations* 1.2.28 (Langella 1999) was used to estimate Nei's pairwise D_A distances (Nei *et al.* 1983), and to construct neighbour-joining unrooted phylograms (Saitou and Nei 1987) based on D_A distances between sample collections. Levels of confidence of phylogenetic groupings were estimated using bootstrapping methods implemented in the program *Populations* 1.2.28, resampling across loci 1000 times. Output was visualized using the program *TreeView* 1.6.6 (Roderic 2001). Factorial correspondence analysis, adapted for molecular genetic data (She *et al.* 1987) and implemented here using the program *Genetix* version 4.02, was carried out to visualize the relative similarity of genotypes of individual salmon from different populations, and to assess the relative similarity of entire sample collections obtained from different rivers. Factorial correspondence analysis of individuals is particularly useful here because it makes no *a priori* assumptions about population membership, providing insight into possible unknown sub-structuring within previously assumed populations. Genetic structuring across Cape Breton samples was also investigated using the model-based clustering program *Structure* (Pritchard *et al.* 2000; Falush *et al.* 2003). The program *Structure* is somewhat unique in that rather than estimating genetic similarity between predefined groups (usually sets of individuals from a given geographical location), the program, loosely speaking, assigns individuals from any sample collection to specific genetic clusters by minimizing overall linkage and Hardy Weinberg disequilibrium. The program *Structure* can be useful in identifying the presence of genetic structuring that may or may not be reflected by the geographic location of collected samples. Twenty replicate runs were carried out for each value of *K* (where *K* is the number of genetic clusters), for *K*=2 to 9, using a burn-in period of 15,000 iterations followed by 100,000 iterations from which estimates were based. Although the lowest Ln P (D) (or log likelihood score) was observed for K=5, results for K=2 to 7 were presented. The admixture model was assumed because of the possibility of migration and gene flow between populations or rivers.

RESULTS AND DISCUSSION

WITHIN-POPULATION GENETIC VARIATION

Overall, within-population genetic variation in ECB Atlantic salmon was slightly lower than that observed in large reference sample collections analyzed here, but greater than that estimated for the single small reference collection (STW) from the iBoF included in this study. Gene diversity, a common measure of genetic variation and approximately equal to the probability that two alleles sampled from a population at random are different, ranged from 0.833 to 0.864 ($\overline{X}$ =0.850) in sample collections from ECB salmon, slightly lower than was observed in the sample collections obtained from the two large reference populations RKR03 (0.876) and NSH00 (0.876) from NB, and generally speaking, reference collections MAB06 (0.862) and MRG01 (0.877) from WCCB (Table 2). Differences in *GD* between the ECB group of sample collections and both NB (p=0.127) and WCCB (p=0.289) reference groups were not significant. Gene

diversity was higher in each of the seven ECB sample collections compared to the sample collection obtained from the small STW reference population (Table 2), though most pairwise comparisons involving the former were not significant. *Allele richness* (*AR*), or the number of different alleles in a collection standardized for differences in sample size, was also lower in the ECB sample collections, ranging from 13.48 to 15.64 (Table 2) ($\overline{X}$ =14.84) relative to the large NB ($\overline{X}$ =15.58) and WCCB ($\overline{X}$ =15.72) reference collections, though neither pair of between-group comparisons was significantly different (p=0.272, and p=0.349, respectively). Allele richness, which is much more sensitive to population bottlenecks than *GD*, was much higher in each of the ECB sample collections compared to the sample collection obtained from the smaller Stewiacke reference population (12.61) (Table 2).

As reported above, within the ECB group of sample collections surveyed, *GD* varied little, especially compared to that observed among SU sample collections (0.676 to 0.845) analyzed at a very similar set of microsatellite loci (O'Reilly *et al.* 2012). Average *GD* for the group of the North Coast of Cape Breton (NCCB) sample collections (NRA0607, NRV06, BAD10) ($\overline{X}$ =0.849) was very similar to the two SECB sample collections (Grand and Inhabitant rivers) analyzed here ($\overline{X}$ =0.852), and the difference between these two groups was not significant (p=0.8070). As reported above, within the ECB group of sample collections surveyed, *AR* varied moderately, however, once again, this range was very much restricted relative to that observed in the set of sample collections analyzed from the SU (7.39-13.32) (O'Reilly *et al.* 2012), though it should be noted that sample collections were rarified to 26 individuals in the latter analysis compared to 39 here. The group of NCCB sample collections exhibited slightly higher levels of *AR* ($\overline{X}$ =15.375) relative to the SECB sample collections ($\overline{X}$ =14.455), though the difference between these two groups was not significant (p=0.272). The Indian Brook sample collection exhibited the lowest *AR* of all ECB samples analyzed here (13.48) (Table 2).

In addition to reducing *GD* and, to a greater extent *AR*, population bottlenecks may also induce correlations in gene frequencies among (linkage disequilibrium) and within (deviations from Hardy-Weinberg expectations) loci (Wang *et al.* 1998). Multiple pairs of microsatellite loci analyzed here departed from linkage equilibrium expectations in four ECB sample collections, NRA0607, IND0607, GRA10, and INH10 (Table 2). Two of these four, NRA0607 and INH10, also exhibited several significant departures from Hardy-Weinberg expectations (Table 2). Interestingly, three of the four sample collections (IND0607, GRA10 and INH10) also exhibited reductions in *AR* relative to other ECB sample collections. Note that all three indicators of population bottleneck effects were observed in the small Stewiacke reference population. Deviations from linkage equilibrium and Hardy-Weinberg expectations were also observed in the large Nashwaak (NSH00) sample collection, though reductions in *GD* and *AR* relative to other sample collections were not.

Taken together, these data indicate that sample collections from SECB, and the single river analyzed from the Bras d'Or Lake (Indian Brook) show possible signs of recent population bottleneck effects and resulting reductions in genetic variation. Once again, reductions in within-population genetic variation were not nearly as great as observed for SU and iBoF sample collections (O'Reilly *et al.* 2012).

Levels of within-population genetic variation observed in ECB salmon may also have been influenced by recent and ongoing gene flow. Although Atlantic salmon are generally believed to exhibit strong spawning site fidelity, tagging studies indicate that a portion of returning adult Atlantic salmon do stray (Jonsson *et al.* 2003). Introgression of new and potentially different genes will result if a) strays encounter and enter a non-native river, b) strays spawn successfully

with native salmon to produce hybrids, and c) the hybrids survive, return, and successfully spawn with native salmon. Strays entering a given river may be from neighbouring populations in adjacent watersheds, or from much more distant locations. Indeed, Dionne *et al.* (2008) found that dispersal among regional groupings was of a similar magnitude as was observed within regional groupings. Given the comparatively large size of nearby Gulf populations (COSEWIC 2010), strays from the Gulf may be important in maintaining within-population genetic variation in coastal ECB populations. The complex coastline of Cape Breton Island and the variable offshore environment (marine versus brackish lacustrine waters) may influence the likelihood of more distant strays from adjacent regions encountering and ascending a given river. In other words, gene flow and the introduction of genetic variation from regional strays might be expected to be restricted in Bras d'Or Lake populations. However, both *GD* and AR were slightly elevated in BAD and MDV populations of Great Bras d'Or lake relative to both NRA and NRV that drain almost directly into the Atlantic Ocean. Both levels of genetic variation were reduced in the IND population of Bras d'Or Lake relative to all four populations. Perhaps Barra Strait, one more partial barrier, may have limited migration of strays from more distant groupings, or from neighboring populations inhabiting rivers that empty in to Great Bras d'Or Lake.

Gene flow mediated by migrants that do encounter and enter rivers of the Bras d'Or lakes, and in particular, Bras d'Or Lake, could also be constrained further by local adaptation, as hypothesized by Dionne *et al.* (2008) for salmon from Quebec, northern NB and Labrador. The watershed occupied by Indian Brook is entirely contained within the Cape Breton Hills Ecodistrict, a unique region characterized by high and low elevation hills forested by tolerant hardwood, spruce and fir. Offspring of migrants from (a) the more distant Gulf region or (b) geographically proximate but ecologically distinct watersheds MDV and BAD (comprised mostly of Cape Breton Highlands), may exhibit poor survival in the IND river, thereby minimizing introgression of new genetic variation through time.

Patterns of within-population genetic variation observed across ECB may also have been influenced by human activities, primarily stocking of populations with salmon produced and reared in captivity. The effects of stocking on patterns of genetic variation can be complex. Within particular rivers, stocking using local broodstock may directly reduce effective population size and levels of genetic variation through what has been referred to as the Ryman-Laikre effect, where the magnitude of reduction is expected to increase along with demographic increases in the combined (hatchery plus wild) population (Ryman and Laikre 1991). However, both the magnitude and direction of effects could vary among programs. Supplementation with local broodstock, where family size of hatchery salmon was equalized at the time of release, has led to slight increases in genetic variation (Hedrick *et al.* 2000). Also, under specific conditions, stocking of multiple families obtained from one or more populations exhibiting allele frequency distributions different from the recipient population could lead to increases in levels of genetic variation. Although such introductions may result in initial increases in levels of within-population genetic variation, later population reductions resulting from fitness effects of either a) non-locally adapted genes or b) domesticated genes, may result in future reductions in population size (Hindar *et al.* 1991) followed by loss of genetic variation via genetic drift. East Cape Breton rivers surveyed here all have extensive histories of stocking, and likely vary in terms of a) when stocking was initiated, b) recency of last stocking events, c) total duration of stocking, d) life stage stocked, e) percent of population taken for broodstock, f) number of families and variance in family size at release, g) initial stocking efficacy (survival and reproductive success of releases), h) subsequent population survival and population growth, and i) broodstock source (local versus non-local). Although empirical studies (Christie *et al.* 2012) indicate that the overall impact of stocking is usually in the direction of reducing levels of genetic variation, it would be difficult to predict the relative effects of the combined variables (a to i) on present-day levels of genetic variation across individual ECB rivers.

AMONG-POPULATION GENETIC STRUCTURING

All pairwise estimates of ECB sample collections analyzed here were significantly different from zero, and ranged in magnitude from 0.0035 to 0.0363, $\overline{X}$ =0.0193 (Table 3). The range and mean F_{ST} values observed were much reduced relative to among-river estimates for SU Atlantic salmon analyzed at a very similar set of microsatellite markers (0.014 to 0.168, $\overline{X}$ =0.054) (O'Reilly *et al.* 2012). Samples collections from NCCB in particular, with the exception of NRV06, exhibited very little genetic differentiation (F_{ST}=0.0035-0.0.0083, $\overline{X}$ F_{ST}=0.0.00653), considerably less than any pair of sample collections obtained from different Maritime rivers surveyed here or by O'Reilly *et al.* (2012) and Vandersteen Tymchuk *et al.* (2010) at a largely overlapping set of marker loci. The sample collection, NRV06, despite the geographical proximity of the associated river to other NCCB rivers, was relatively differentiated from all other NCCB and SECB sample collections analyzed here (Table 3), and will be discussed further below. IND0607, INH10, and GRA10 were moderately differentiated from each other and all remaining ECB sample collections (Table 3); levels of structuring were similar to or slightly less than that observed among sample collections obtained from the larger more genetically variable populations of the Northeast SU grouping (those corresponding to SFA 20). Expanding analyses to include sample collections from the WCCB of the adjacent DU, mean pairwise F_{ST} of Cape Breton sample collections actually declined to 0.0170. This reduced among-population differentiation is mostly driven by low levels of genetic structuring between MRG from the WCCB and NCCB sample collections (excluding NRV06) (F_{ST}<0.0086). These results and levels of differentiation observed between RKR03 and MRG01 (F_{ST} >0.01) may not be entirely consistent with the existing DU structure proposed for the region (COSEWIC 2010).

Results from the phylogenetic analyses based on Nei's D_A distance (Figure 2) are largely concordant with patterns of structuring described above based on pairwise F_{ST} values. The NCCB sample collections, excluding NRV06, first join together before grouping with the MRG01-NRV06, and then MAB06-INH10 clusters. IND06 is highly divergent from all other sample collections, and joins the phylogeny between the large group of Cape Breton sample collections, and the remaining cluster of sample collections obtained from other regions or DUs. GRA10 actually clusters first with SME07 from the adjacent SU DU, before joining the NSH00-RKR03-STW01 group. Again, the clustering of WCCB sample collections with ECB sample collections, rather than RKR03 from their assigned DU, and GRA10 with SME07 of the SU, is not entirely consistent with existing DU structure delineated in COSEWIC (2010). It should be noted that a) statistical support of many branch points is relatively week, b) many ECB pairs join at a point not very distant from more major branch nodes, and c) that distances between these major nodes, particularly in that part of the phylogeny that includes Cape Breton sample collections, are limited. This again reflects the overall modest level of structuring observed among most Cape Breton sample collections, including between WCCB and many from ECB.

In the factorial correspondence analyses of the nine sample collections from Cape Breton Island, IND0607 is again highly divergent from all other sample collections, separating out on primary axis 1 (accounting for 20.5 percent of the variance), and from all other sample collections except NRV on axis 3 (Figure 3a). As in the analyses discussed above, MDV06 and BAD10 cluster together before grouping with NRA06 and MRG01, and then MAB06. After removing the highly divergent IND0607 sample collection in a second factorial correspondence analysis of sample collections, the relatively close grouping NCCB (less NRV06) and WCCB sample collections is again evident, as is the divergence of INH10, GRA10 and NRV06 from all other sample collections, including each other (Figure 3b). Factorial correspondence analyses of individuals from these same populations demonstrates that individual samples from the latter

three populations cluster separately from all other populations, but that there is considerable spatial overlap of individuals from the remaining populations (Figure 4).

In the cluster analysis carried out by the program *Structure*, nearly all individuals from NRA0607, BAD10, MDV06, GRA10, and MRG01, and approximately 1/4 to 1/2 of those from NRV06, INH10 and MAB06 appear to originate from multiple (*K*) inferred genetic clusters, in roughly similar proportions (Figure 5, panels K=3 to 7). This is the pattern expected if very little structuring exists in the collection of samples analyzed (Pritchard *et al.* 2000). A large portion (approaching 100%) of all IND0607 salmon, however, appear to originate from a single genetic cluster (see green vertical bars, Figure 5, panel K=5). Note that although this can be seen in K=5-7, this pattern is most apparent in the analysis specifying *K*=5 genetic clusters, the value of *K* associated with the lowest Ln P (D) (or log likelihood score) of all the values of *K* tested. The partitioning of individual genomes from nearly all IND0607 samples into a single genetic cluster with very limited representation in any other sample collection, is consistent with results from the phylogenetic and factorial correspondence analyses discussed above, suggesting that IND0607 is genetically distinct from all other sample collections. Note too that a large proportion of the genomes of a large percentage of NRV06 samples (approximately 1/2 of the samples analyzed) originated from a second genetic cluster (Figure 5, panel K=5 (red), panel K=6 (green), panel K=7 (yellow); no such individuals are observed in any other sample collection analyzed. These results suggest the possible presence of two genetically distinct groups of Atlantic salmon within the NRV06 collection of samples analyzed. These could correspond to two groups reproductively isolated in space (spawning in different tributaries), time (spawning in early versus late fall) or through some unknown behavioural mechanism. The presence of this divergent sub-group identified in the *Structure* analysis, may have driven the phylogenetic and factorial correspondence results above, where NRV06 is somewhat divergent from all other NCCB sample collections.

The possible role of undetected kinship in contributing to *Structure* results observed in this population was investigated using the program Colony (Wang 2004), which rules of uses Mendelian inheritance, resulting expected patterns of distributions of alleles in groups of full and half siblings (sib), and allele frequency information to identify putative full- and half-sib offspring groups. No full sib group of three or more, and no half-sib group of five or more individuals, were detected, and most individuals (>90%) were sole representatives of different full sib groups. These results indicate that the different genetic clusters observed in NRV06 were not due to the existence of one or more large kin groups in this sample collection.

Finally, it should also be noted that a large portion of the genomes of a substantial subset of individuals form INH10 and MAB06 appear to originate from a third genetic cluster, most apparent in the *K*=5 *Structure* analysis. In the absence of geographic information, these results could be interpreted as indicating the presence of strays or migrants (or their early generation offspring) from one river in another. However, it is difficult to reconcile this observation with the geographic juxtapositioning of Mabou and Inhabitants rivers, the date of construction of the Canso Causeway and the permanent closing of the Strait of Canso (1954). Still, it is interesting to note that similarity between MAB06 and INH10 was also observed in the phylogenetic analyses discussed above, and that prior to the causeway, these river mouths were separated by a few tens of kilometres of coastline. Although seemingly unlikely, the observed patterns could be explained by pre-causeway gene flow between locations, and maintenance of two reproductively isolated sub-groups of similar salmon in both rivers. A hypothetical example would be pre-causeway establishment of genetically similar early run salmon in both rivers, and maintenance of reproductive isolation from genetically divergent late run salmon in both (and possibly other nearby) rivers.

IMPLICATIONS OF OBSERVED WITHIN- AND AMONG-POPULATION GENETIC VARIATION FOR THE PRIORITIZATION OF CONSERVATION EFFORTS FOR EAST CAPE BRETON ATLANTIC SALMON

All salmon populations from a given region potentially contribute genetically or demographically to the long-term persistence of a given DU and possibly the species itself and are therefore important. However, resources for biological conservation are sometimes scarce and undesirable decisions, often of great risk, must be made. A number of different approaches have been suggested for prioritizing species or other taxonomic groups for conservation, recently discussed in O'Reilly (2006) and O'Reilly and Doyle (2007). Ultimately, decisions would ideally be based on many criteria including both a) molecular genetic and genetically based phenotypic differences in quantitative traits (Crandall *et al.* 2000), and b) ecological and life history information (Utter *et al.* 1993). Here, results are presented only on analyses of neutral molecular genetic data, though it is recognized that this is only part of the picture.

As mentioned earlier, the approach for prioritizing populations for conservation developed by Petit *et al.* (1998) considers both within- and among-population genetic variation. Use of within-population genetic variation in prioritizing populations for conservation is relatively straightforward; genetic variation is maximized by prioritizing populations exhibiting higher levels of within-population genetic variation, specifically allele richness. Among-population genetic variation is maximized by selecting populations for prioritization (or culling populations not to be conserved) that minimize the overall tree length in the remaining group of conserved populations. The rationale behind this approach is explained in Figure 6. In principle, populations or species that are less phylogentically divergent (that cluster closer together in a phylogenetic tree) are expected to exhibit greater overlap in hypothetical character traits that have a genetic basis. Selection or prioritization of taxonomic groups B and C at the expense of taxonomic group A (Figure 6a), results in much reduced overall tree length associated with the remaining populations (B and C), and a small set of different character traits retained (see also text in Figure 6). On the other hand, selection or prioritization of taxonomic groups A and C at the expense of taxonomic group B (Figure 6b), results in an overall tree length little reduced from the original, and much longer than the tree in Figure 6a. This alternate selection of populations (A and C) results in the retention of a much larger set of different character traits (see also text in Figure 6).

Given the existing DU framework (ignoring similarity between ECCB sample collections and those from the WCCB, and similarity between GRA10 and SU sample collections), and, for the moment, considering among-population genetic variation only, genetic variation would be maximized by conserving one of the three NRA, BAD, MDV populations from the main NCCB grouping as well as the relatively divergent IND population from the Bras d'Or Lake. Based on pairwise F_{ST} values and Nei's D_A distances (and the associated phylogenetic tree), less among-population genetic variation would be lost by next prioritizing the relatively divergent GRA or INH populations. Based on among-population criteria once again, NRV would be prioritized next, followed by one of the remaining two non-NCCB populations not originally prioritized. Less weight could be given to within-population genetic variation because ECB sample collections included in this study varied little in *AR* (13.48 to 15.64, and 14.16 to 15.64 if IND0607 is excluded). Consideration of *AR* following prioritization based on among-population variation would result in the prioritization of MDV (*AR*=15.64) over BAD (*AR*=15.44) and NRA (*AR*=15.38) in the initial selection and prioritization of one of these three phylogenetically similar rivers. The slightly higher levels of *AR* for the sample collection obtained from GRA relative to INH would warrant prioritizing of the former over the latter. Consideration of within-versus among-genetic variation is in conflict in prioritization of IND; F_{ST} values and the phylogenetic analyses suggest IND should be prioritized high, yet low levels of *AR* (13.48) suggest it should be prioritized last.

Again, because populations varied little in *AR*, more weight could be given to among-population information in this instance. Finally, earlier it was indicated that NRV should be prioritized after GRA and INH, and before NRA and BAD. However, as noted previously, it is possible that the NRV consists of two genetically divergent groups of individuals, and that the more divergent group contributed to both the *Structure* analysis results and the phylogenetic placement of the entire sample collection. This possibly distinct genetic cluster (represented by the red vertical lines in Figure 5, panel 5), particularly if found to be concordant with some hypothetical phenotypic trait differences (*e.g.* adult return date, early summer versus fall), could be important to conserve.

CAVEATS AND LIMITATIONS

Given the objectives of this study (identification and prioritization of within- and among-population biodiversity), the sample collections analyzed here are not ideal. First, only a small portion of ECB rivers that could potentially harbour Atlantic salmon were sampled. Although it is hoped that information obtained from one population may be somewhat representative of others from that area (*e.g.* Indian Brook and other Bras d'Or Lake populations), potentially important differences can exist between geographically proximate rivers (*e.g.* NRV and NRA) that cannot always be predicted by geographic distance, suspected barriers to gene flow, or phenotypic differences. Second, sample sizes obtained from individual locations were restricted, impacting the accuracy of estimates of both within- and among-population genetic variation. Indeed, error in estimating pairwise F_{ST} is 1/2N, where N is the sample size obtained from a given location (Waples 1998). Given the shallow structuring observed here, this is particularly noteworthy. Third, sample collections consisted exclusively of juveniles. Limited dispersal of fry and parr from redds, combined with geographically restricted sampling, can result in collections of salmon being comprised of relatively few kin groups, and overall sample collections that poorly reflect the actual population. More specifically, the presence of family structuring (multiple half- or full-sibs) within sample collections can inflate estimates of between-population differences, and result in an inappropriate finding of statistically significant differences between populations that may be homogeneous (see Allendorf and Phelps 1981). It should be noted, however, that sample sizes here were limited by the low abundance of juveniles encountered in most of the sampled rivers, and limited resources for sample collection.

Apart from the above sample limitations, other conditions encountered here were sub-optimal for the model-based *Structure* analysis, which could have led to spurious conclusions. First, *Structure* analyses, particularly "unsupervised" analyses where population information is not used to help assign individuals, does not perform well when levels of structuring (though potentially real and statistically significant) are weak. Levels of population structuring observed here, inferred from traditional FST analyses was limited (0.0035 to 0.0363). Second, because the program *Structure* is based on the optimal allocation of genomes to genetic clusters so as to minimize linkage and Hardy-Weinberg dis-equilibrium, the true populations must be in linkage and Hardy-Weinberg equilibrium. Some true ECB populations may have deviated from linkage and Hardy-Weinberg equilibrium, possibly because of historic or recent bottleneck effects. Additionally, several results produced by *Structure* indicate that the program was having difficulty identifying genetic clusters, including a) temporally variable alpha values, and b) multi modal values of ΔK when attempting to infer true *K*.

Finally, there exists considerable controversy over the correspondence between patterns of putatively neutral molecular genetic variation and patterns of variation in genetically based phenotypic differences of adaptive significance, the real target of conservation measures. Usually, very little information is available on the existence and distribution of actual local adaptive differences among-populations being considered for prioritization. To mitigate these

concerns, it is important to base conservation decisions on multiple sources of relevant information, including neutral genetic marker data, ecological information, phenotypic information, tagging data, and demographic information.

ACKNOWLEDGEMENTS

We thank Kris Gordos, Andrea Cox, Louise de Mestral, Manon Cassista Da-Ros, and Roxanne Gillett for assisting with laboratory analyses. Ross Jones, Heather Bowlby, Leroy Anderson, Peter Amiro, and Beth Lenentine contributed samples analyzed in this study. Carolyn Harvie helped check microsatellite genotype data for errors. Lorraine Hamilton reviewed later drafts of this manuscript, for which we are also appreciative. Funding for this research was provided by Fisheries and Oceans Canada's Species at Risk Program, and the Canadian Regulatory System for Biotechnology.

REFERENCES

Allendorf, F.W., and Phelps, S.R. 1981. Use of allelic frequencies to describe population structure. Canadian Journal of Fisheries and Aquatic Sciences 38: 1507-1514.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., and Bonhomme, F. 2000. GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UPR 9060, Université de Montpellier II, Montpellier (France).

Christie, M.R., Marine, M.L., French, R.A., Waples, R.S., and Blouin, M.S. 2012. Effective size of a wild salmonid population is greatly reduced by hatchery supplementation. Heredity 109: 254-260.

COSEWIC 2010. COSEWIC assessment and status report on the Atlantic Salmon *Salmo salar* (Nunavik population, Labrador population, Northeast Newfoundland population, South Newfoundland population; Southwest Newfoundland population, Northwest Newfoundland population, Quebec Eastern North Shore population, Quebec Western North Shore population, Anticosti Island population, Inner St. Lawrence population, Lake Ontario population, Gaspé-Southern Gulf of St. Lawrence population, Eastern Cape Breton population, Nova Scotia Southern Upland population, Inner Bay of Fundy population, Outer Bay of Fundy population) in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. xlvii + 136 pp. (Last accessed 3 September 2013).

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M., and Wayne, R.K. 2000. Considering evolutionary processes in conservation biology. Trends in Ecology and Evolution 15: 290-295.

DFO and MNRF. 2008. Conservation Status Report, Atlantic Salmon in Atlantic Canada and Quebec: PART I - Species Information. Canadian Manuscript Report of Fisheries and Aquatic Sciences 2861. 208 p.

Dionne, M., Miller, K.M., Dodson, J.J., Caron, F., and Bernatzhez, L. 2007. Clinal variation in MHC diversity with temperature: Evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. Evolution 61(9): 2154-2164.

Dionne, M., Caron, F., Dodson, J.J., and Bernatchez, L. 2008. Landscape genetics and hierarchical genetic structure in Atlantic salmon: The interaction of gene flow and local adaptation. Molecular Ecology 17: 2382-2396.

Falush, D., Stephens, M., and Pritchard, J.K. 2003. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Goudet, J. 1995. FSTAT (version 1.2): A computer program to calculate F-statistics. Journal of Heredity 52: 275-278.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). (Last accessed 12 August 2013).

Hedrick, P.W., Hedgecock, D., Hamelberg, S., and Croci, S.J. 2000. The impact of supplementation in winter-run Chinook salmon on effective population size. Journal of Heredity 91(2):112-116.

Hindar, K., Ryman, N., and Utter, F. 1991. Genetic effects of cultured fish on natural fish populations. Canadian Journal of Fisheries and Aquatic Sciences 48: 945-957.

Hurlbert, S. H. 1971. The nonconcept of species diversity: A critique and alternative parameters. Ecology 52: 577-586.

Jonsson, B., Jonsson, N., and Hansen, L.P. 2003. Atlantic salmon straying from the River Imsa. Journal of Fish Biology 62: 641-657.

King, T., Eackles, M.S., and Letcher, B.H. 2005. Microsatellite DNA markers for the study of Atlantic salmon *(Salmo salar)* kinship, population structure, and mixed-fishery analyses. Molecular Ecology Notes 5: 130-132.

Langella, O. 1999. Populations 1.2.28. Available from [Email] Olivier.langella@moulon.inra.fr. (Last accessed 3 September 2013).

MacMillan, J.L., Caissie, D., LeBlanc, J.E., and Crandlemere, T.J. 2005. Characterization of summer water temperatures for 312 selected sites in Nova Scotia. Canadian Technical Report of Fisheries and Aquatic Sciences 2582. 43 p.

Mohseni, O., and H.G. Stefan. 1999. Stream temperature/air temperature relationship: A physical interpretation. Journal of Hydrology 218: 128-141.

Nei, M., Tajima, F., and Tateno, Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. Journal of Molecular Evolution 19: 153-170.

Neily, P. D., Quigley, E., Benjamin, L., Stewart, B., and Duke, T. 2003. Ecological land classification for Nova Scotia Volume 1 – Mapping Nova Scotia's terrestrial ecosystems. Nova Scotia Department of Natural Resources Renewable Resources Branch. Report DNR 2003-2. 83 p.

Nevo, E. 1978. Genetic variation in natural populations: Patterns and theory. Theoretical Population Biology 13: 121-177.

Nicieza, A.G., Reiriz, L., and Brana, F. 1994. Variation in digestive performance between geographically disjunct populations of Atlantic salmon: countergradient passage time and digestion rate. Oecologia 99: 243-251.

O'Reilly, P.T. 2006. Towards the identification of Conservation Units in Atlantic salmon from Eastern Canada. Department of Fisheries and Oceans Canadian Science Advisory Secretariat Research Document 2006/012. vi + 39 p.

O'Reilly, P., and Doyle, R. 2007. Live Gene Banking of endangered populations of Atlantic Salmon. *In* The Atlantic salmon: Genetics, conservation and management. Edited by E. Verspoor, L. Stradmeyer, and J.L. Nielsen. Blackwell Publishing, Oxford, UK. pp. 436-482.

O'Reilly, P., Hamilton, L., McConnell, S. K., and Wright, J. M. 1996. Rapid analysis of genetic variation in Atlantic salmon *(Salmo salar)* by PCR multiplexing of tetranucleotide and

dinucleotide microsatellites. Canadian Journal of Fisheries and Aquatic Science 53: 2292-2298.

O'Reilly, P., Rafferty, S., and Gibson, A.J.F. 2012. Within- and among-population genetic variation in the Southern Upland Designatable unit of Maritime Atlantic salmon (*Salmo salar* L.) Fisheries and Oceans Canada. Canadian Science Advisory Secretariat. Research Document. 2012/077. iv + 17 p.

Parker, M., Westhead, M., Doherty, P., and Naug, J. 2007. Ecosystem Overview and Assessment Report for the Bras d'Or Lakes, Nova Scotia. Canadian Manuscript Report of Fisheries and Aquatic Sciences 2789. xxii + 223 p.

Paterson, S., Piertney, S.B., Knox, D., Gilbey J., and Verspoor, E. 2004. Characterisation and PCR multiplexing of novel highly variable tetranucleotide Atlantic salmon *(Salmo salar)* microsatellites. Molecular Ecology Notes 4: 160-162.

Petit, R.J., Mousadik, A., and Pons, O. 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12: 844–855.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Riddell, B.E., and Leggett, W.C. 1981. Evidence of an adaptive basis for geographical variation in body morphology and time of downstream migration of juvenile Atlantic salmon *(Salmo salar)*. Canadian Journal of Fisheries and Aquatic Sciences 38:308-320.

Roderic, D.M. 2001. TreeView 1.6.6. (Last accessed 3 September 2013).

Rousset, F. 2008. GENEPOP'007: A complete re-implementation of the Genepop software for Windows and Linux. Molecular Ecology Resources 8: 103-106.

Ryman, N., and Laikre, L. 1991. Effects of supportive breeding and variance effective population size. Conservation Biology 9: 1619-1628.

Saccheri, I., Kuussaari, M., Kankare, P.V., Fortelius, W., and Kanski, I. 1998. Inbreeding and extinction in a butterfly metapopulation. Nature 392: 491-494.

Saitou, N., and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4(4): 406-425.

She, J.X., Autem, M., Kotulas, G., Pasteur, N., and Bonhomme, F. 1987. Multivariate analysis of genetic exchanges between *Solea aegyptiaca* and *Solea senegalensis* (Teleosts, Soleidae). Biological Journal Linnean Society 32: 357–371.

Slettan, A., Olsaker, I., and Lie, O. 1995. Atlantic salmon, *Salmo salar*, microsatellites at the SSOSL25, SSOSL85, SSOSL311, SSOSL417 loci. Animal Genetics 26: 281–282.

Taylor, E.B. 1991. A review of local adaptation in Salmonidae, with particular reference to Pacific and Atlantic salmon. Aquaculture 98: 185-207.

Utter, F., Seeb, J.E., and Seeb, L.W. 1993. Complementary uses of ecological and biochemical genetic data in identifying and conserving salmon populations. Fisheries Research 18: 59–76.

Vandersteen Tymchuk, W., O'Reilly, P.T., Bittman, J., Macdonald, D., and Schulte, P. 2010. Conservation genomics of Atlantic salmon: Variation in gene expression between and within regions of the Bay of Fundy. Molecular Ecology 19: 1842-1859.

Verspoor, E., and Jordan, W.C. 1989. Genetic variation at the Me-2 locus in the Atlantic salmon within and between rivers: Evidence for its selective maintenance. Journal of Fish Biology 35: 205-213.

Verspoor, E., O'Sullivan, M., Arnold, A.L., Knox, D., and Amiro, P.G. 2002. Restricted matrilineal gene flow and regional differentiation among Atlantic salmon (*Salmo salar* L.) populations within the Bay of Fundy, Eastern Canada. Heredity 89: 465–472.

Wang, J. 2004. Sibship reconstruction from genetic data with typing errors. Genetics 166: 1963-1979.

Wang, J., Caballero, A., and Hill, W.G. 1998. The effect of linkage disequilibrium and deviation from Hardy-Weinberg proportions on the changes in genetic variance with bottlenecking. Heredity 81(2): 174-186.

Waples, R.S. 1991. Pacific salmon, *Oncorhynchus spp.*, and the definition of "species" under the Endangered Species Act. Maritime Fisheries Review 53: 11-22.

Waples, R.S. 1998. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. Journal of Heredity 89: 438-450.

Weir, B.S., and Cockerham, C.C. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

TABLES

Table 1. Sample collection information, including river name and code, collection year, location, Designatable Unit, Salmon Fishing area, and sample size.

Sample Collection Name	Sample Collection Code	Collection Year	Latitude/Longitude	DU	Salmon Fishing Area	Sample Size
North River Aspy-1	NRA0607	2006,2007	46-54-16;60-30-50	13-ECB	19	44
North River-2	NRV06	2006	46-18-15;60-37-10	13-ECB	19	73
Baddeck River-3	BAD10	2010	46-06-02; 60-50-23	13-ECB	19	52
Middle River-4	MDV06	2006	46-05-06;60-54-37	13-ECB	19	73
Indian Brook-5	IND0607	2006,2007	45-56-37;60-36-11	13-ECB	19	52
Grand River-6	GRA10	2010	45-38-52;60-39-41	13-ECB	19	53
Inhabitants River-7	INH10	2010	45-38-53;61-13-23	13-ECB	19	53
Mabou River-8	MAB06	2006	46-04-04; 61-23-02	12-GSG	18	80
Margaree River-9	MRG01	2001	46-24-27;61-04-42	12-GSG	18	49
Kedgwick River-10	RKR03	2003	47-39-51; 67-29-29	12-GSG	16	58
St. Mary's East Branch-11	SME07	2007	45-15-22; 62-03-49	14-SU	20	59
Stewiacke River-12	STW01	2001	45-08-27; 63-23-34	15-iBoF	22	82
Nashwaak River-13	NSH00	2000	45-57-22; 66-37-18	16-BoF	23	70

Note: ECB=East Cape Breton; GSG=Gaspé-Southern Gulf of St. Lawrence; SU=Southern Upland; iBoF=inner Bay of Fundy; oBoF=outer Bay of Fundy. Latitude/Longitude=Latitude North; Longitude West (degrees-minutes-seconds). DU-Designatable unit (COSEWIC 2010).

Table 2. Measures of within-population genetic variation and F_{IS} (inbreeding coefficient) for East Cape Breton and several other reference sample collections.

	NRA0607	NRV06	BAD10	MDV06	IND0607	GRA10	INH10	MAB06	MRG01	RKR03	smE07	STW01	NSH00
GD avg	0.855	0.833	0.854	0.852	0.852	0.864	0.840	0.862	0.877	0.876	0.839	0.811	0.876
GD var	0.017	0.021	0.017	0.018	0.010	0.008	0.011	0.010	0.008	0.007	0.027	0.015	0.007
ObH avg	0.858	0.831	0.849	0.859	0.869	0.873	0.835	0.839	0.894	0.864	0.844	0.812	0.848
ObH var	0.020	0.027	0.023	0.020	0.010	0.008	0.016	0.013	0.009	0.008	0.028	0.015	0.004
Ar^{39d} avg	15.38	15.04	15.44	15.64	13.48	14.75	14.16	15.35	16.10	15.56	15.07	12.61	15.60
Ar var	41.58	36.68	77	51.58	22.73	44.29	37.64	40.92	43.95	33.86	46.09	31.10	42.06
NaO avg	15.67	17.27	16.33	17.27	14.13	15.67	15.20	17.47	16.87	16.93	16.20	14.87	17.27
NaO var	42.81	52.21	49.95	71.07	26.84	53.81	48.03	57.41	52.41	42.78	58.03	46.70	55.78
LD (p<0.05)	21	5	6	5	18	14	31	12	7	4	13	40	29
LD (p<0.01)	14	3	4	2	9	7	21	7	4	2	8	26	20
HWE (p<0.05)	3	0	2	1	2	0	4	0	2	1	0	4	6
HWE (p<0.01)	2	0	1	0	0	0	2	0	1	0	0	3	4
F_{IS} *(<0)*	8	12	8	8	10	9	9	3	10	4	9	8	3
F_{IS} *(<-0.05)*	4	1	2	0	4	2	0	1	4	0	1	3	1
F_{IS} *(<-0.075)*	1	0	1	0	3	0	0	0	1	0	0	1	1
F_{IS} *(>0)*	8	4	8	8	6	7	7	13	6	12	7	8	13
F_{IS} *(>0.05)*	2	3	2	1	0	0	3	5	1	1	1	3	3
F_{IS} *(>0.075)*	1	3	1	0	0	0	1	2	0	0	0	2	2

Note: *GD* avg. = Average *Gene Diversity*; *GD* var. = cross locus variance in *Gene Diversity*; *ObH* avg. = Average *Observed Heterozygosity*; *ObH* var. = cross locus variance in *Observed Heterozygosity*; Ar^{39d} avg. = Average *Allele Richness* standardized to 39 diploid individuals; *Ar* var. = cross locus variance in *Allele Richness*; *NaO* avg. = Average observed number of alleles; *NaO* var. = cross locus variance in *Observed number of alleles*; *LD (p<0.05)* = number of pairs of loci that deviate from linkage equilibrium at alpha 0.05 or less; *LD (p<0.01)* = number of pairs of loci that deviate from linkage equilibrium at alpha 0.01 or less; *HWE (p<0.05)* = number of loci that deviate from Hardy Weinberg expectations at alpha 0.05 or less; *HWE (p<0.01)* = number loci that deviate from Hardy Weinberg expectations at alpha 0.01 or less; F_{IS} *(<0)* = number of single locus F_{IS} values less than 0; F_{IS} *(<-0.05)* = number of single locus F_{IS} values less than -0.05; F_{IS} *(<-0.075)* = number of single locus F_{IS} values less than -0.075; F_{IS} *(>0)* = number of single locus F_{IS} values greater than 0; F_{IS} *(>0.05)* = number of single locus F_{IS} values greater than 0.05; F_{IS} *(>0.075)* = number of single locus F_{IS} values greater than 0.075. Full sample names corresponding to alphanumeric codes are given in Table 1.

Table 3. Pairwise F_{ST} (θ) estimates of between sample collections obtained from Cape Breton and reference locations in the years 2001-2010.

	NRV06	BAD10	MDV06	IND0607	GRA10	INH10	MAB06	MRG01	RKR03	smE07	STW01	NSH00
NRA0607	0.0229	0.0083	0.0078	0.0201	0.0182	0.0197	0.0175	0.0077	0.0140	0.0421	0.0473	0.0244
NRV06	-	0.0229	0.0184	0.0339	0.0170	0.0363	0.0234	0.0173	0.0257	0.0334	0.0604	0.0348
BAD10	-		0.0035	0.0165	0.0184	0.0159	0.0136	0.0096	0.0140	0.0428	0.0412	0.0256
MDV06	-	-	-	0.0156	0.0185	0.0134	0.0110	0.0084	0.0126	0.0408	0.0414	0.0253
IND0607	-	-	-	-	0.0271	0.0265	0.0217	0.0172	0.0205	0.0508	0.0476	0.0290
GRA10	-	-	-	-	-	0.0242	0.0155	0.0093	0.0141	0.0175	0.0468	0.0168
INH10	-	-	-	-	-	-	0.0124	0.0136	0.0168	0.0428	0.0458	0.0271
MAB06	-			-	-	-	-	0.0105	0.0124	0.0264	0.0405	0.0174
MRG01	-	-	-	-		-	-	-	0.0103	0.0287	0.0433	0.0150
RKR03	-	-	-	-	-	-	-	-	-	0.0314	0.0361	0.0099
smE07	-	-	-	-	-	-	-	-	-	-	0.0673	0.0248
STW01	-	-	-	-	-	-	-	-		-	-	0.0466

Note: All pairwise estimates significant at $p<0.05$; full sample names corresponding to alphanumeric sample codes are given in Table 1.

FIGURES

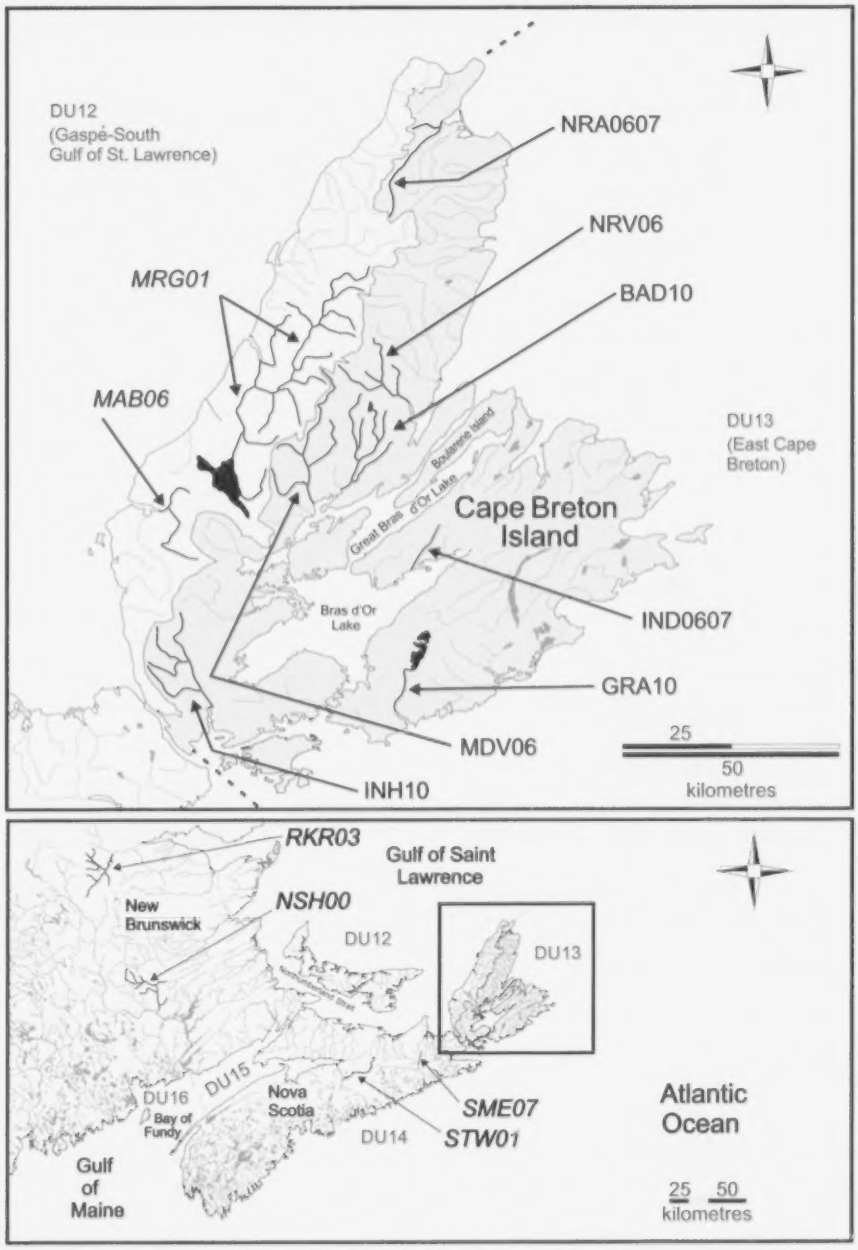

Figure 1. Geographic location of sampled rivers from the East Cape Breton Designatable Unit (top panel, grey background) and reference rivers from throughout Maritimes Canada (top and bottom panels). Full sample names corresponding to alphanumeric codes are given in Table 1.

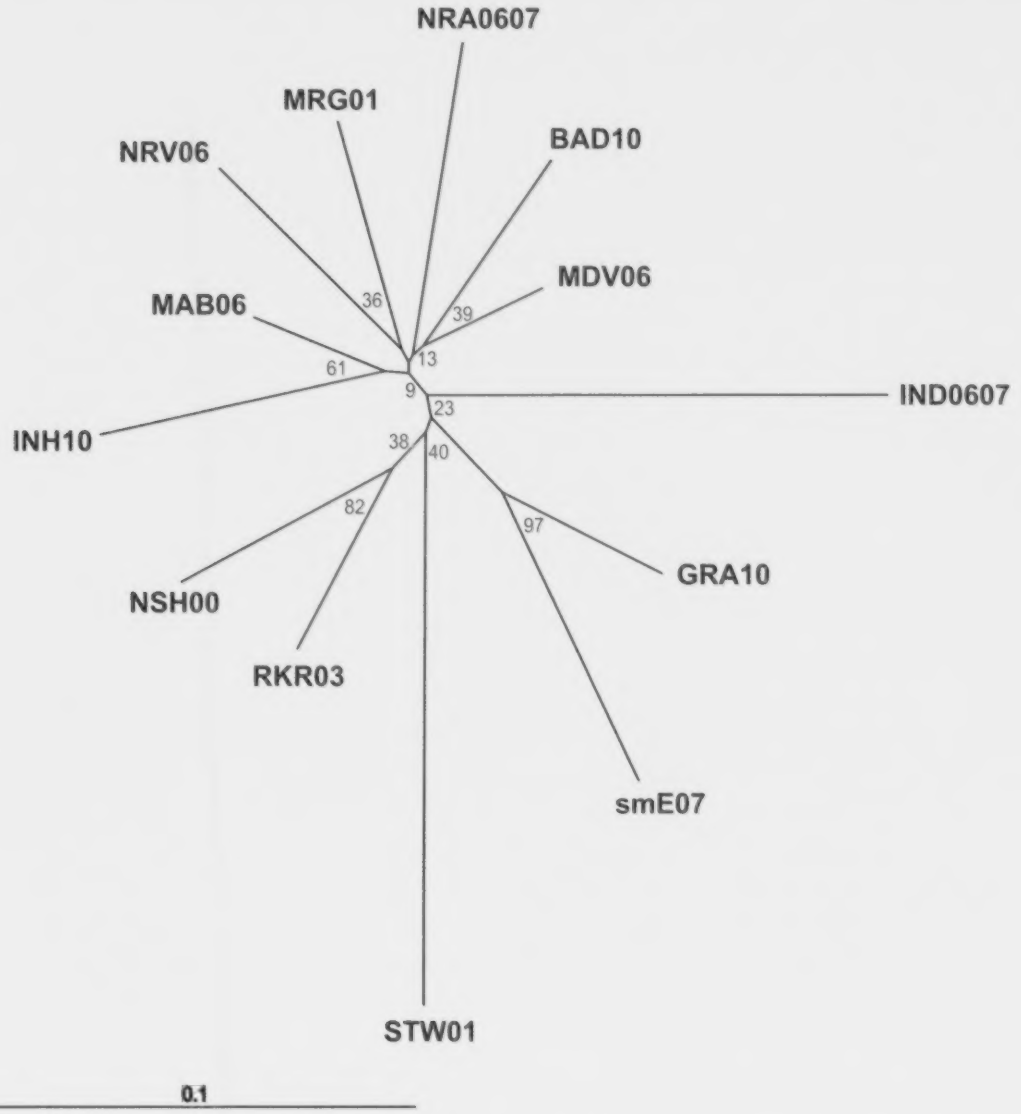

Figure 2. Population phylogeny based on Nei's D_A pairwise genetic distances, constructed using the neighbour-joining method. Numbers near branch nodes indicate level of bootstrap support obtained by resampling across loci (with replacement) 1000 times. Full sample names corresponding to alphanumeric sample codes are given in Table 1.

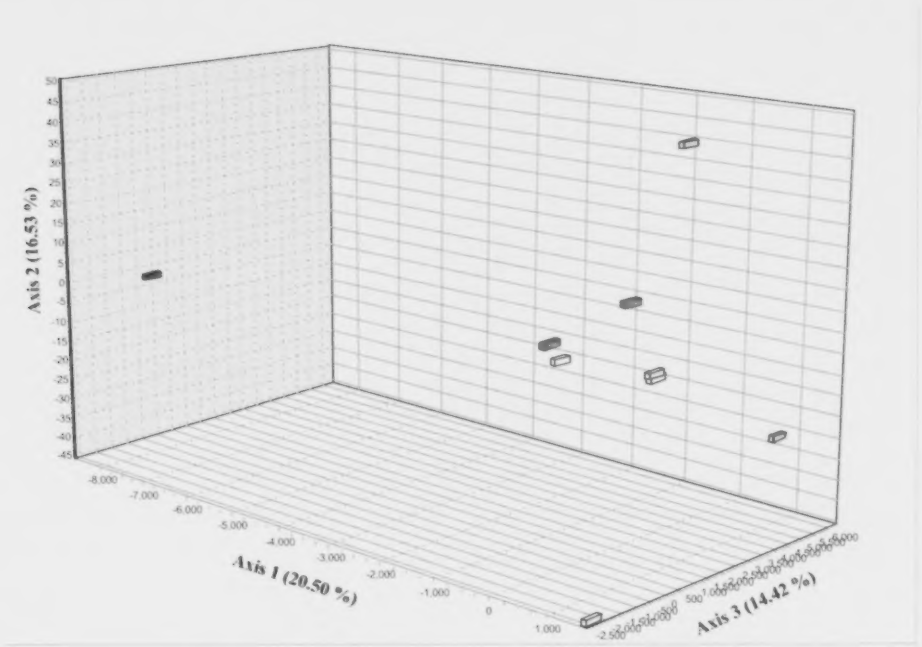

Figure 3a. Factorial correspondence analysis of multilocus microsatellite genotype information for all nine populations obtained from Cape Breton. Population of origin is identified by colour (see key above). Full sample names corresponding to alphanumeric sample codes are given in Table 1.

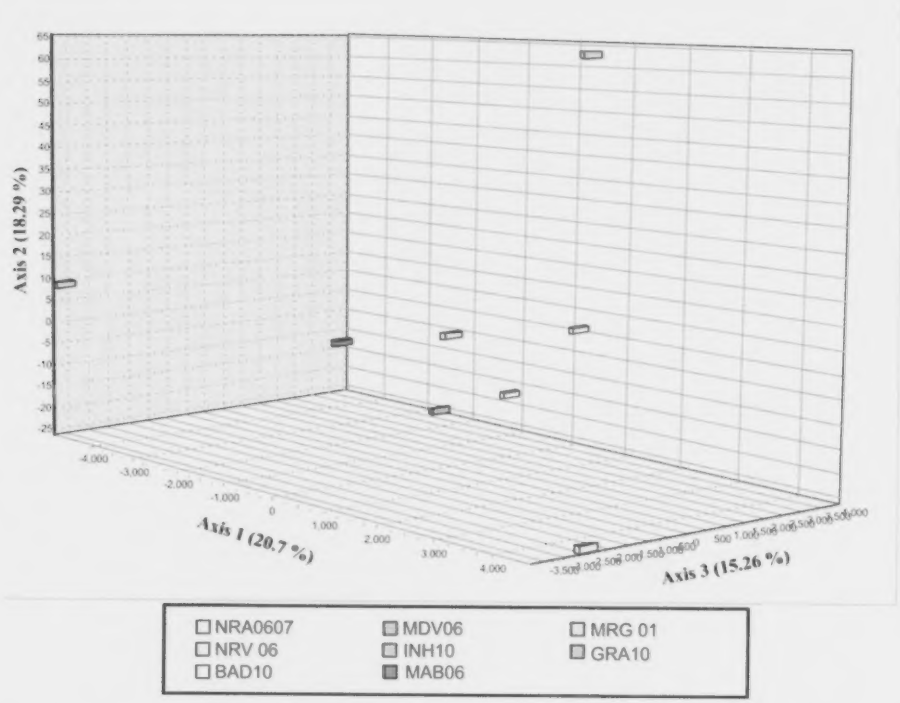

Figure 3b. Factorial correspondence analysis of multilocus microsatellite genotype information for eight of nine populations obtained from Cape Breton. Population of origin is identified by colour (see key above). Full sample names corresponding to alphanumeric sample codes are given in Table 1.

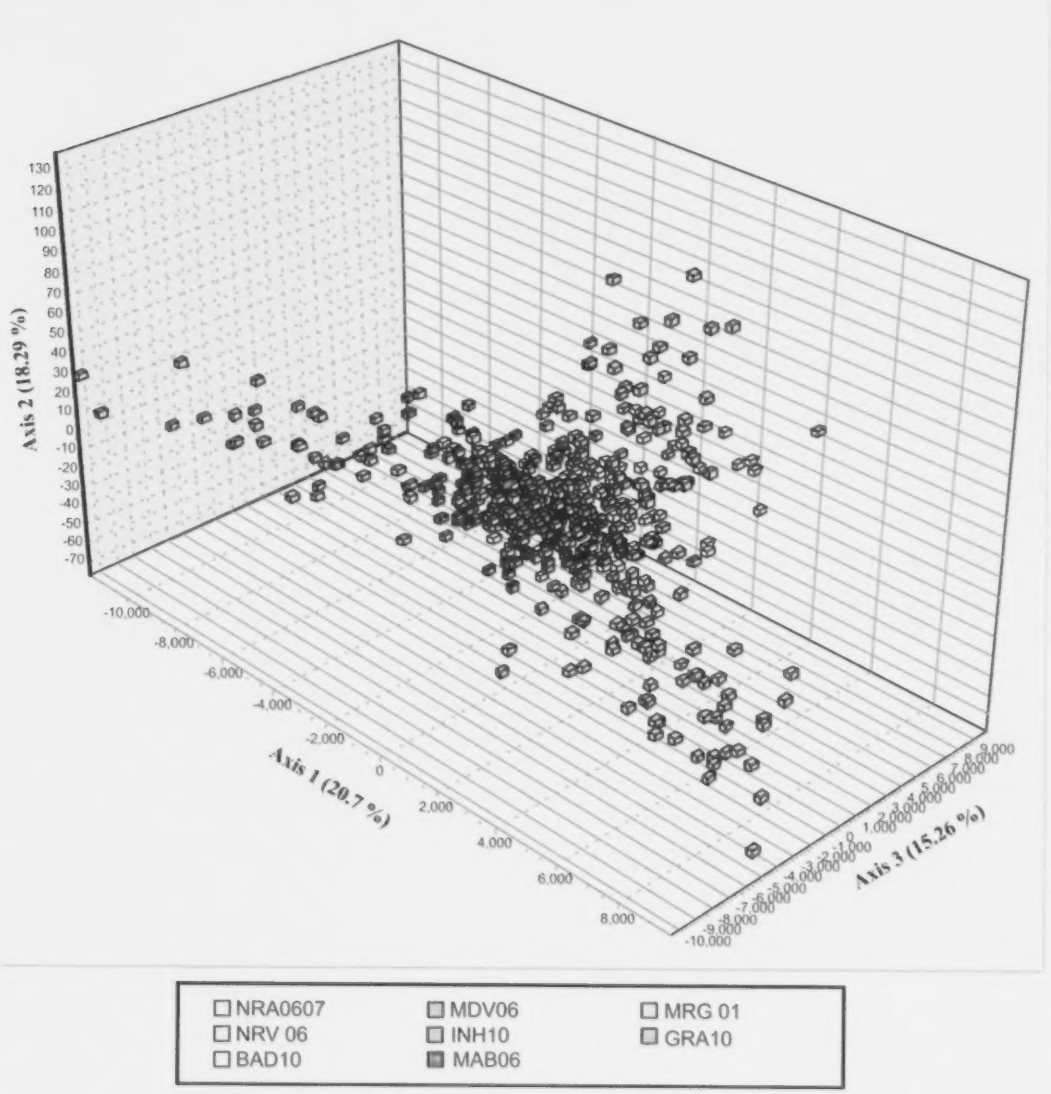

Figure 4. Factorial correspondence analysis of multilocus microsatellite genotype information for individuals from eight populations obtained from Cape Breton (the divergent IND0607 was excluded to increase separation of remaining individuals). Population of origin is identified by colour (see key above). Full sample names corresponding to alphanumeric sample codes are given in Table 1.

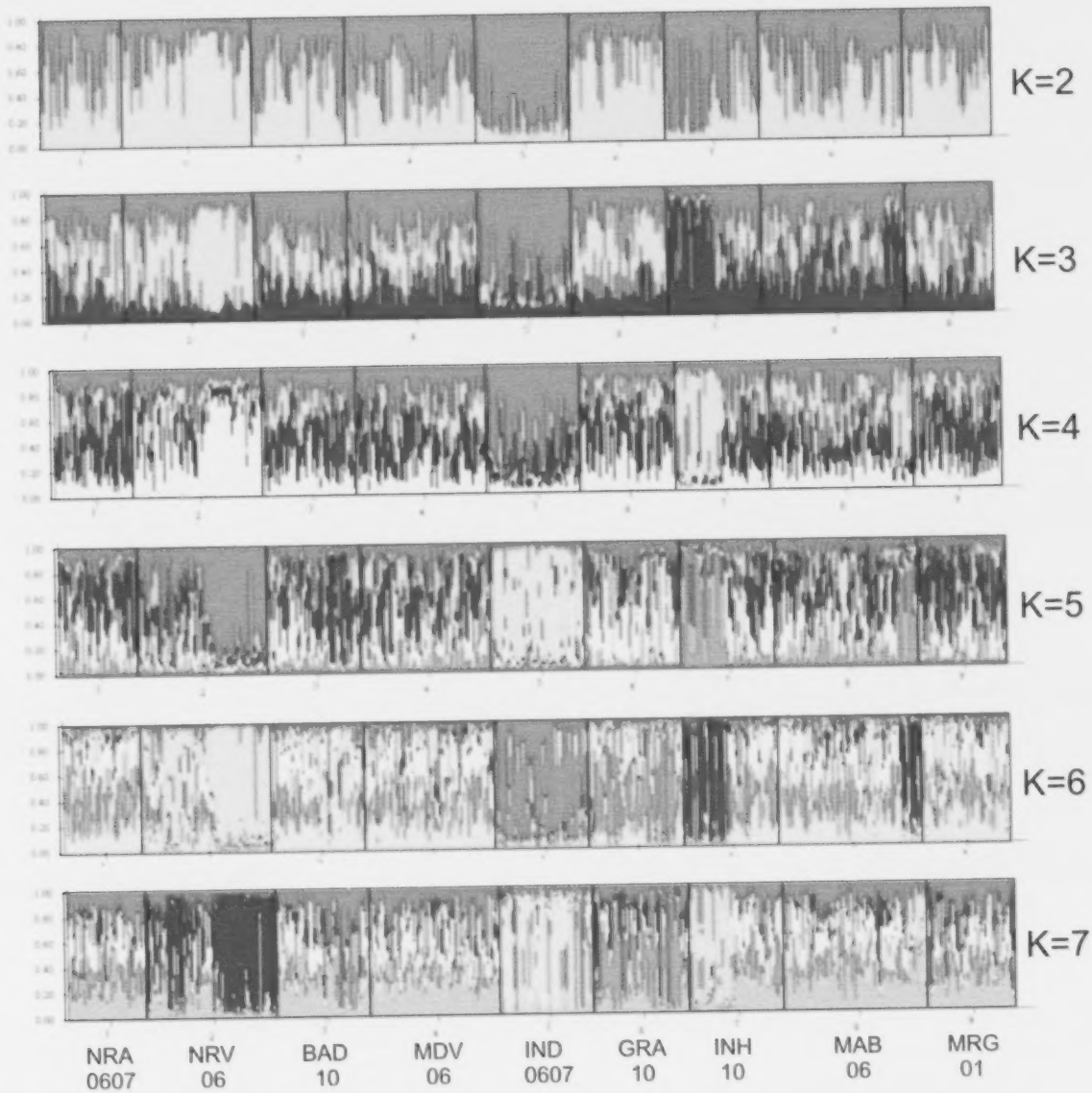

Figure 5. Partitioning of individual genomes (vertical lines) into genetic clusters, where clusters are denoted by colour. The number of genetic clustered specific to a given analysis ranges from 2 to 7 (right side of figure). Individuals are arranged by population to facilitate comparisons.

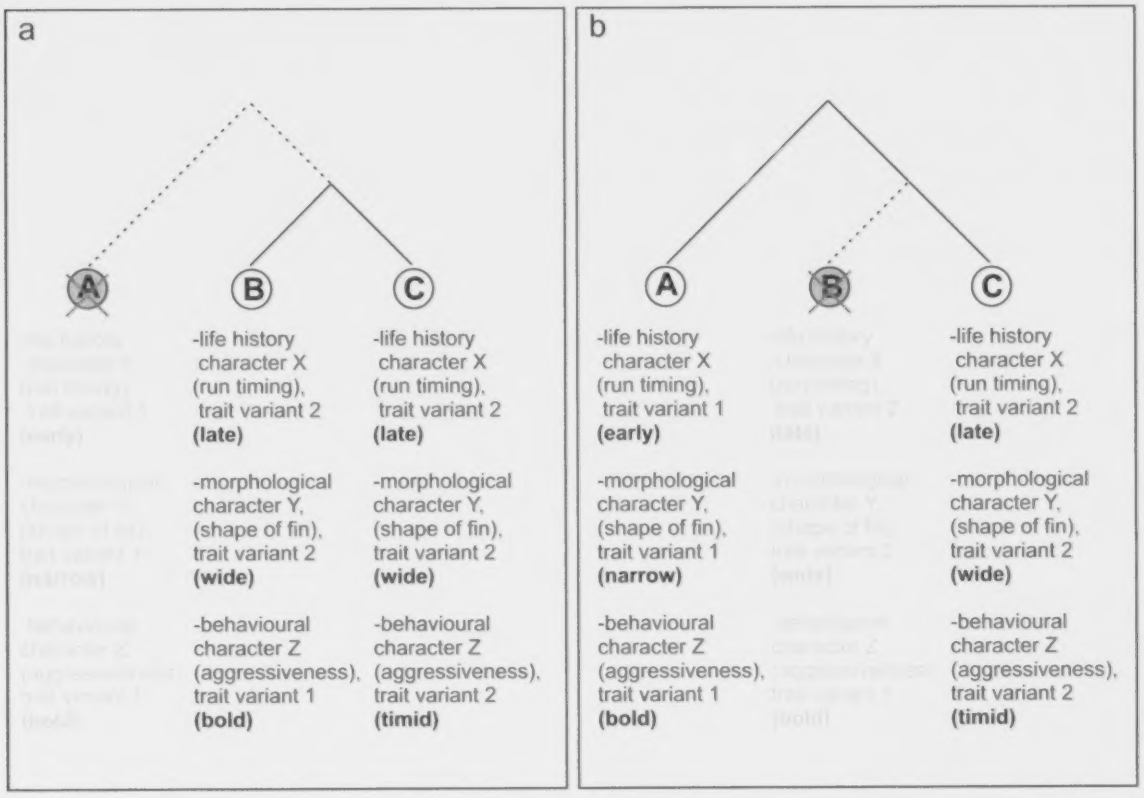
a
A
B
C
-life history character X (run timing), trait variant 2 (late)
-morphological character Y, (shape of fin), trait variant 2 (wide)
-behavioural character Z (aggressiveness), trait variant 1 (bold)
-life history character X (run timing), trait variant 2 (late)
-morphological character Y, (shape of fin), trait variant 2 (wide)
-behavioural character Z (aggressiveness), trait variant 2 (timid)
b
A
B
C
-life history character X (run timing), trait variant 1 (early)
-morphological character Y, (shape of fin), trait variant 1 (narrow)
-behavioural character Z (aggressiveness), trait variant 1 (bold)
-life history character X (run timing), trait variant 2 (late)
-morphological character Y, (shape of fin), trait variant 2 (wide)
-behavioural character Z (aggressiveness), trait variant 2 (timid)

Figure 6. Phylogeny of hypothetical taxonomic groups A, B and C that vary with respects to hypothetical life history, morphological, and behavioural characteristics. Loss of group A (left panel) results in a greater reduction in overall tree length, and a greater net loss of potentially important biological variation compared to loss of group B (right panel).